



The Complete Genome of the Uncultivated Ultra-Deep Subsurface Bacterium *Desulforudis audaxviator* Obtained by Environmental Genomics

D. Chivian^{1,2 *}, E.J. Alm^{1,3}, E.L. Brodie², D.E. Culley⁴, T.Z. DeSantis², T.M. Gihring⁵, A. Lapidus⁶, L-H. Lin⁷, S. Lowry⁶, D.P. Moser⁸, P. Richardson⁶, G. Southam⁹, G. Wanger⁹, L.M. Pratt¹⁰, G.L. Andersen², **A.P. Arkin**^{1,2,11,12}, T.C. Hazen^{1,2}, F.J. Brockman⁴, T.C. Onstott¹³
^{*}*DCChivian@lbl.gov*, ¹*VIMSS*, ²*BNL*, ³*MIT*, ⁴*PNNL*, ⁵*FSU*, ⁶*JGI*, ⁷*NTU, Taiwan*, ⁸*DRI*, ⁹*U. Waterloo, Canada*, ¹⁰*Indiana*, ¹¹*U.C. Berkeley*, ¹²*HHMI*, ¹³*Princeton*



ABSTRACT

"*In Sneffels Joculis craterem quem delibat Umbra Scartaris Julii intra calendas descende, **audax viator**, et terrestre centrum attinges.*" ("Descend, **bold traveller**, into the crater of the jokul of Sneffels, which the shadow of Scartaris touches before the kalends of July, and you will attain the center of the earth.")
-- Hidden message deciphered from an Icelandic saga that prompts Professor Lidenbrock to undertake his journey in Jules Verne's "Journey to the Center of the Earth."

A more complete picture of life on Earth, and even life *in* the Earth, has recently become possible through the application of environmental genomics. We have obtained the complete genome sequence of a new genus of the *Firmicutes*, the uncultivated sulfate reducing bacterium *Desulforudis audaxviator*, by filtering fracture water from a borehole at 2.8 km depth in a South African gold mine. The DNA was sequenced at the JGI using a combination of Sanger sequencing and 454 pyrosequencing, and assembled into just one genome, indicating the planktonic community is extremely low in diversity. We analyzed the genome of *D. audaxviator* using the MicrobesOnline annotation pipeline and toolkit (<http://www.microbesonline.org/>), which offers powerful resources for comparative genome analysis, including operon predictions and tree-based comparative genome browsing. MicrobesOnline allowed us to compare the *D. audaxviator* genome with other sequenced members of the *Firmicutes* in the same clade (primarily *Pelotomaculum thermopropionicum*, *Desulfotomaculum reducens*, *Carboxydotherrnus hydrogenoformans*, and *Moorella thermoacetica*), as well as other known sulfate reducers (including *Archaeoglobus fulgidus* and *Desulfovibrio vulgaris*). *D. audaxviator* gives a view to the set of tools necessary for what appears to be a self-contained, independent lifestyle deep in the Earth's crust. The genome is not very streamlined, and indicates a motile, endospore forming sulfate reducer with pili that can fix its own nitrogen and carbon. *D. audaxviator* is an obligate anaerobe, and lacks obvious homologs of many of the traditional O₂ tolerance genes, consistent with the low concentration of O₂ in the fracture water and its long-term isolation from the surface. *D. audaxviator* is valuable in that it provides a complete genome representative of the Gram-positive bacteria to further our understanding of dissimilatory sulfate reducing bacteria and archaea. Additionally, study of the deep subsurface has offered access to the simplest community yet studied by environmental genomics, perhaps consisting of just a single organism that may be capable of performing all of the tasks necessary for life.

METHODS

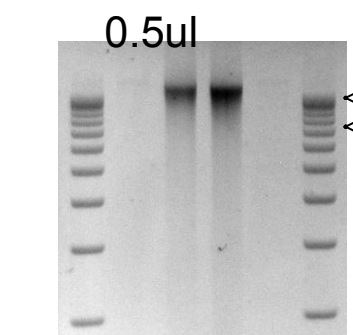
Collection of sample

Borehole starting from 2.8 kmbls accessing fracture water ranging down to 4-5 kmbls. Cells density was ~3.3x10⁴ cells/mL and ~5.6x10⁶ mL passed across 0.5 µm filter -> ~1.8x10¹¹ cells.



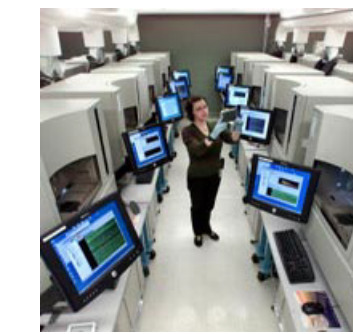
Extraction of DNA from filters

3 extractions with DNAzol yielding 46 µg of HMW DNA.



Sequencing

Sanger reads: 20.7 Mb of high quality (29,279 reads with average trimmed length of 708 bp).
454 reads: 56.2 Mb (518,272 reads with average length of 109 bp).



Genome annotation

Used MicrobesOnline (<http://www.microbesonline.org>) for ortholog and protein family assignment (e.g. COG and Pfam), operon prediction, and tree-based genome browser for comparative analysis.

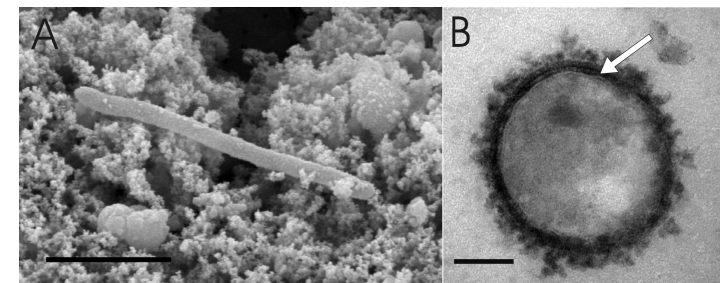


GENERAL FEATURES

Feature	Value
Genome size, bp	2,349,476
G+C content, %	60.9
Predicted protein coding genes (CDS or ORF)	2157
Genes without homology to other organisms (ORFans)	210
Pseudogenes derived from a protein coding gene	82
Average CDS length, bp	910
Longest CDS length, bp	5601
Percent of genome protein coding, %	86.8
Non-coding mutations in high quality Sanger reads	16
Synonymous mutations in high quality Sanger reads	22
Non-synonymous mutations in high quality Sanger reads	29
Ribosomal RNA operons (16S-23s-5S)	2
Transfer RNAs (all AA, incl. Sec)	45
Structural RNAs	4
CRISPR regions	2
Transposon gene groups	30
Transposon genes	83
Integrases	17

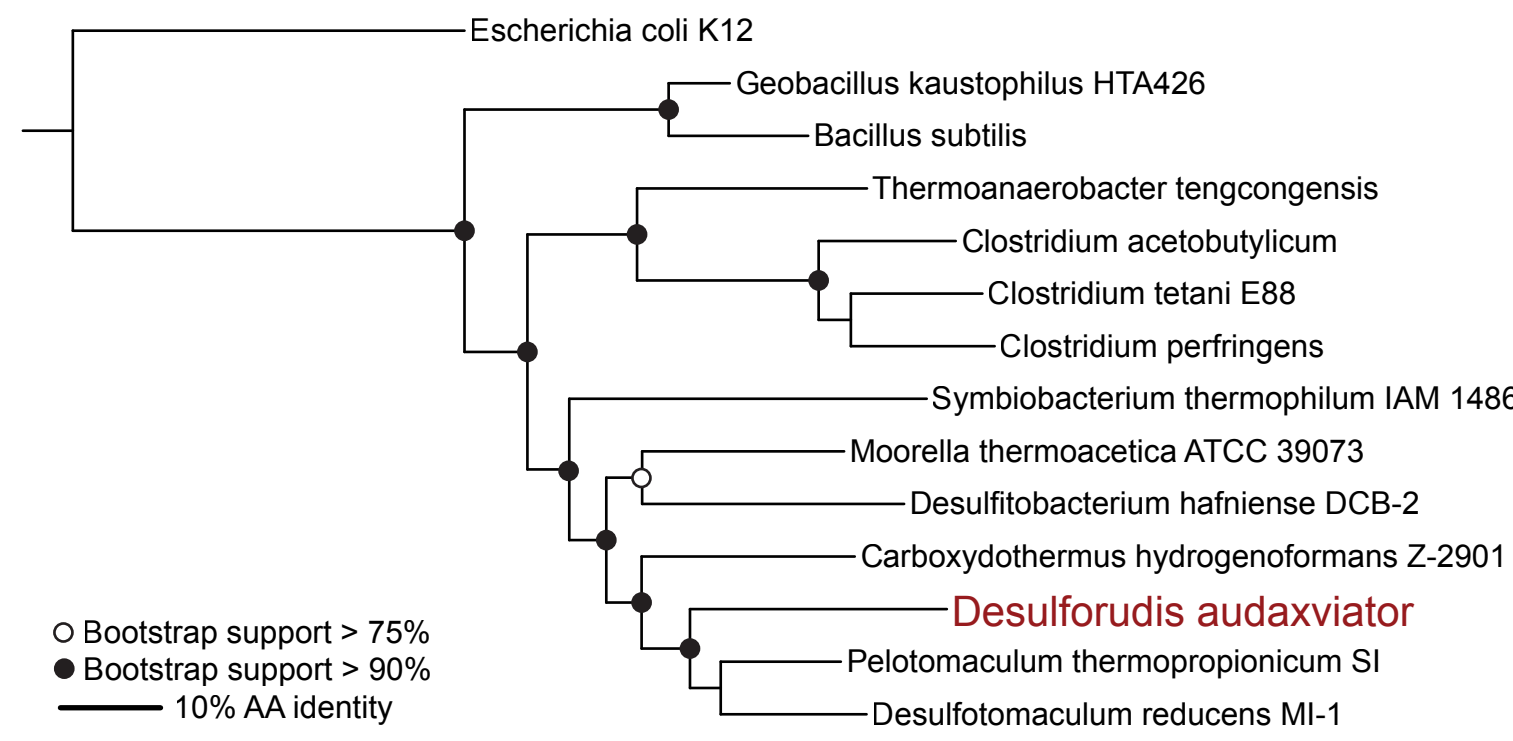
IMAGING

(A) SEM and (B) TEM micrographs of dominant microorganism from 648 meters down in borehole D8A of Moser et al. (Moser, Gihring et al. 2005) where *D. audaxviator* is the predominant microorganism. Bars equal 2 µm and 100 nm respectively, arrow indicates cell envelope.



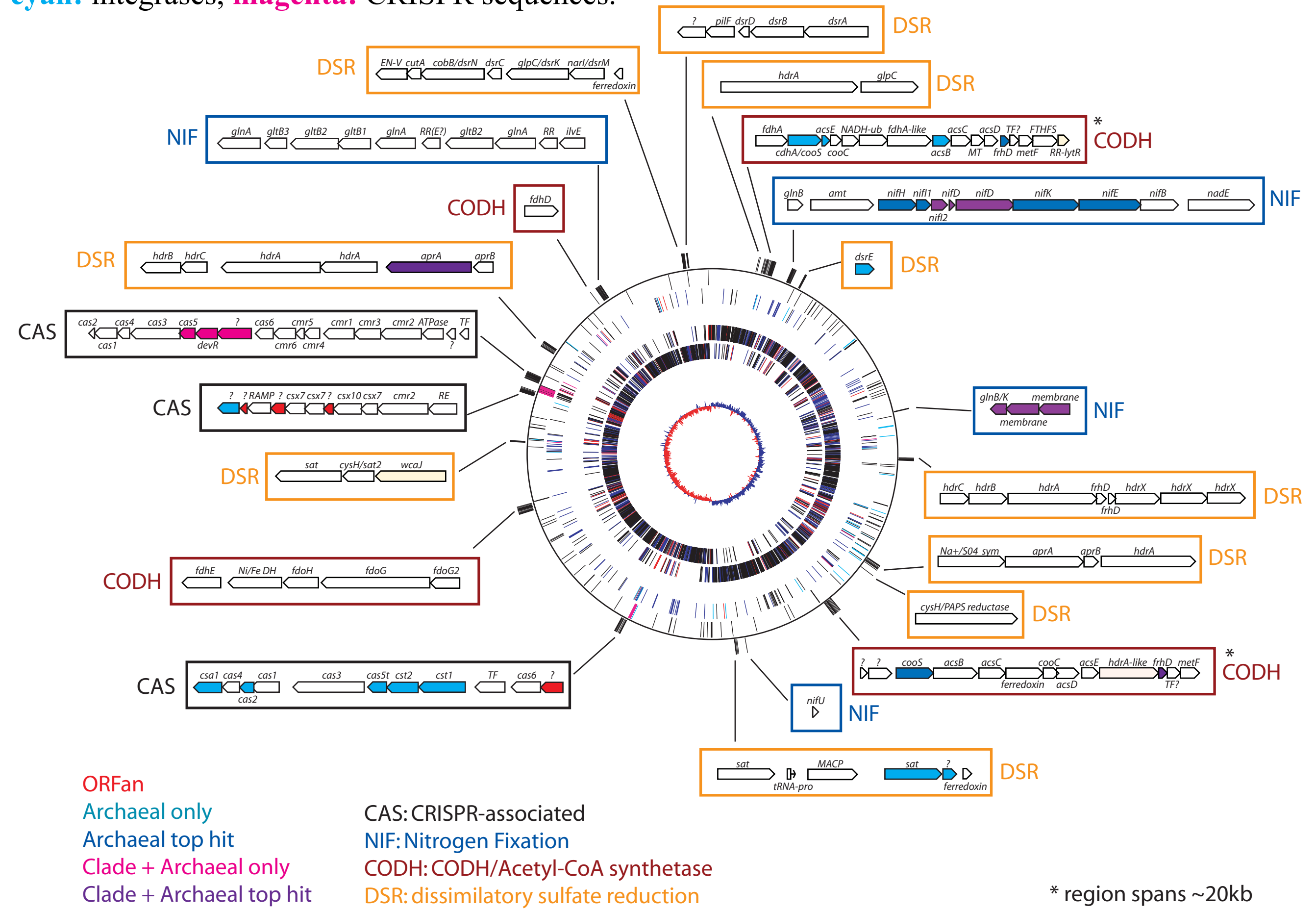
PHYLOGENY

Phylogenetic tree based on protein sequences of univeraly distributed clusters of orthologous groups (Ciccarelli, Doerks et al. 2006). Sequences aligned with MUSCLE (Edgar 2004). Tree determined by maximum likelihood with PHYML (Guindon and Gascuel 2003) using JTT substitution model (Jones, Taylor et al. 1992).

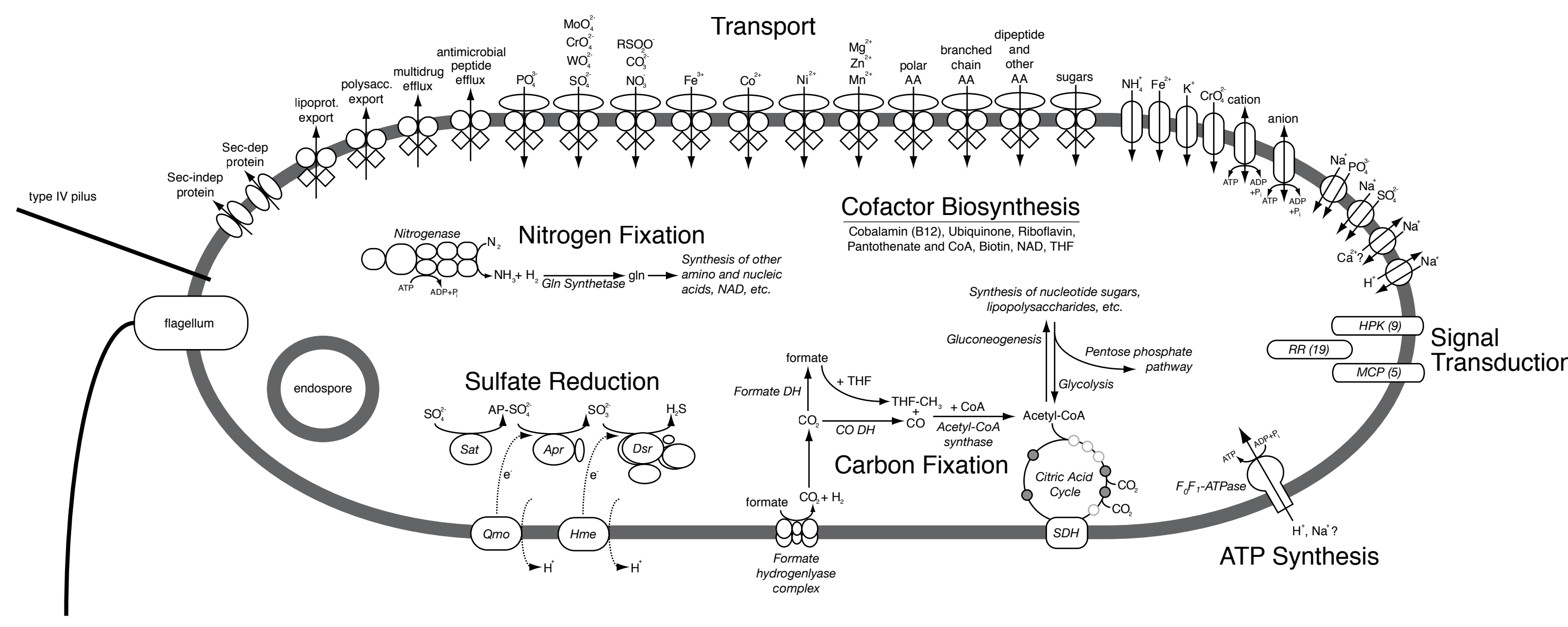


GENOME

Innermost ring: G+C skew. **Second and third rings:** protein coding genes on each strand, with **black:** genes with strongest homology to genes found in closest clade species, **blue:** genes with strongest homology to genes in species not in closest clade, **red:** genes without homology to genes found in any other organism (ORFans). **Fourth ring:** classification of clade-specific and archaeal associated genes, with **black:** genes found only in nearest clade, **cyan:** genes with homologs found only in archaea, **blue:** genes for which the closest hit is archaeal, **magenta:** genes with hits (including the top) to members of the immediate clade and archaea only, **purple:** genes with hits (including the top) to archaea and members of the immediate clade. **Fifth ring:** locations of mobile and repetitive elements, with **black:** transposons, **cyan:** integrases, **magenta:** CRISPR sequences.



PHYSIOLOGY



HORIZONTAL GENE TRANSFER

D. audaxviator possess many genes found in Archaea and only rarely in bacteria or only within the immediate clade of the *Firmicutes*, including many thought to be indicative of hyperthermophiles and thermophiles (Makarova, Wolf, Koonin 2003). Additionally, several of the key genes involved in nitrogen and carbon fixation, as well as CRISPR associated genes, are shared with Archaea.

Gene ID	Name	Description	AA ID	Species of closest homolog
1357786	fbp	COG1980 Archaeal fructose 1,6-bisphosphatase	79.12	Pelotomaculum thermopropionicum SI
1358995		COG1042 Acyl-CoA synthetase (Succinyl-CoA syn.?)	62.46	Pelotomaculum thermopropionicum SI
1358298	nifH	COG1348 Nitrogenase subunit NifH (ATPase)	79.27	Methanothermobacter thermautotrophicus
1358299	nifI	COG347 Nitrogen regulatory protein PII	64.00	Methanococcus maripaludis
1358300	nifD	COG347 Nitrogen regulatory protein PII	51.64	Methanosarcina acetivorans C2A
1358302	nifD	COG2710 Nitrogenase molybdenum-iron protein, alpha	67.86	Moorella thermoacetica ATCC 39073
1358303	nifK	COG2710 Nitrogenase molybdenum-iron protein, beta	46.55	Methanococcus maripaludis
1358304	nifE	COG2710 Nitrogenase molybdenum-iron protein	38.86	Methanothermobacter thermautotrophicus
1358618	glnB2	PF00543 Nitrogen regulatory protein P-II (GlnB, GlnK)	54.21	Methanospirillum hungatei JF-1
1358619		PF07556 DUF1538 (Putative membrane protein)	58.75	Carboxydotherrnus hydrogenoformans
1358620		PF07556 DUF1538 (Putative membrane protein)	55.46	Carboxydotherrnus hydrogenoformans
1358260	cdhA/cooS1	COG1152 CO DH/acetyl-CoA synthase, alpha	52.37	Methanothermobacter thermautotrophicus
1358261	cdhB/acsE	COG1880 CO DH/acetyl-CoA synthase, epsilon	38.18	Methanococcus maripaludis
1358266	cdhC/aseB	COG1614 CO DH/acetyl-CoA synthase, beta	58.52	Methanocaldococcus jannaschii
1358270	frhD1	COG1908 Coenzyme F420-reducing hydrogenase, delta	44.81	Moorella thermoacetica ATCC 39073
1358970	cooS2	COG1152 CO DH/acetyl-CoA synthase, alpha	51.26	Methanosarcina mazei Goe1
1358978		COG1908 Coenzyme F420-reducing hydrogenase, delta	53.24	Moorella thermoacetica ATCC 39073
1357268	csa1	COG4343 Conserved in archaea	37.83	Archaeoglobus fulgidus
1357269	cas4	COG1468 RecB family exonuclease	34.76	Archaeoglobus fulgidus
1357270	cas2	COG1343 Predicted to be involved in DNA repair	36.49	Methanosarcina mazei Goe1
1357271	cas1	COG1518 Predicted to be involved in DNA repair	40.07	Archaeoglobus fulgidus
1357272	cas3	COG1203 Predicted helicases	31.23	Methanosarcina barkeri str. fusaro
1357273	csf2	COG1857 Predicted to be involved in DNA repair	42.48	Methanosarcina acetivorans C2A
1357274	csf1	[IPR010180: CRISPR-associated, CXXC_CXXC region]	22.74	Methanosarcina barkeri str. fusaro
1357764	cas5	[IPR013422: CRISPR-associated protein Cas5]	63.27	Thermoanaerobacter tengcongensis
1357765	devR	COG1857 Predicted to be involved in DNA repair	75.00	Thermoanaerobacter tengcongensis
1357766		hypothetical protein	65.20	Thermoanaerobacter tengcongensis

DISCUSSION

The extreme dominance of *D. audaxviator* in the fracture water (> 93% of 16S clones, which we believe makes this the most uniform community yet studied) has permitted the assembly of its complete genome from an environmental sample. Analysis of its genome revealed the likely capacity for an independent lifestyle, including nitrogen fixation using machinery shared with archaea and carbon fixation from formate and carbon monoxide via the Acetyl-CoA (Wood-Ljungdahl) pathway. Consistent with the high sulfate concentrations (1860 µM), possibly the result of radiolysis of H₂O, the genome also possess the machinery for dissimilatory sulfate reduction, which exports protons for use by F₀F₁-ATP synthase. The extremely low concentration of O₂ (below the detectable limit of 1 µM) has permitted the loss of all oxygen tolerance genes except for superoxide dismutase and rubrerythrin, consistent with *D. audaxviator*'s long-term isolation (the fracture water is > 3 million years old). However, this streamlining appears to be an exception, and the genome of *D. audaxviator* is otherwise not particularly streamlined like those formed under extreme competition for limited nutrients (e.g. *Prochlorococcus* or *Pelagibacter*), contrary to what might have been expected for a deep subsurface environment. Its 2157 protein coding genes, while on the low-end of the spectrum, is still a typical number for free-living bacteria. Additionally, counter to the belief that an organism that has been in a stable non-competitive environment for a long time will drop unnecessary DNA to reduce its reproductive burden, it also contains a large number of mobile DNA elements (including 30 types of transposons integrated at 83 sites). It also contains CRISPR sequences (and CRISPR-associated proteins), potentially to combat phage invasion. Finally, the presence of genes necessary for sensing and flagellar motion indicates that *D. audaxviator* is likely capable of energetically costly motility. Whether *D. audaxviator* survives as spores during 'famine' stages, emerging periodically for short periods of growth, or migrates along fracture surfaces seeking more favorable environments, remains an important question for further analysis. Geochemical analyses of the source fracture water which revealed that the free energy flux per cell is much higher than required for simple maintenance (Lin, Wang et al. 2006) and that the deep subsurface environments in the South African crust are energy rich (Kieft, McCuddy et al. 2005) compared to shallow subsurface environments. It appears that *D. audaxviator* lives in an even more dynamic environment than had been anticipated, and it remains an open question why it alone has been so successful in colonizing the energy-rich deep subsurface of South Africa, and perhaps elsewhere in the world.

ACKNOWLEDGEMENTS

ESPP is part of the Virtual Institute for Microbial Stress and Survival supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics Program:GTL through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy. Funding was also provided by NASA to the IPTAI Astrobiology Institute (<http://www.indiana.edu/~deeplife>) for the Detection of bio-sustainable energy and nutrient cycling in the deep subsurface of Earth and Mars project..